Optimal Resource Allocation Technique (ORAT) for Green Cloud Computing

K. L. Giridas
Computer Science Engineering Department,
Noorul Islam University,
Kumaracoil, Nagercoil,
Tamilnadu-629 180.

Shajin Nargunam
Computer Science Engineering Department,
Noorul Islam University,
Kumaracoil, Nagercoil,
Tamilnadu- 629 180.

ABSTRACT
As the IT trade progress towards game-changing expertise, a cloud Eco-system is gradually increasing in the country with expertise corporation ramping up employing and guiding for cloud computing. An accomplishment of green IT is probable to assist an organization in several ways like operating cost, stakeholder value, sustainability, employee morale and so on. Cloud computing might hoist privacy and security concerns but this could have one obvious advantage, far enhanced energy efficiency. The previous work studied about the accumulation of e-waste by integrating the old and mid-range processors with modern processors but the care about resource efficiency is less. In this paper, proposes an optimal resource allocation method for cloud computing environments. This paper progress a resource allocation representation of green cloud computing environments, considering both bandwidth and processing capability, allocated concurrently to every service request and returned it on an hourly basis. The owed resources are committed to every service request. It is established by simulation evaluation that the proposed ORAT method can diminish the request loss possibility and therefore, decrease the total resource obligatory, compared with the predictable allocation method. Through the optimal resource allocation, the resources for the tasks are allocated for cloud computing environment by eradicating e-waste and make IT as Green IT. The proposed Optimal resource allocation technique (ORAT) for cloud computing in Green IT can be implemented in CloudSim software, and various performance characteristics can be simulated to estimate the performance of the proposed ORAT in terms of processing ability, resource utilization, bandwidth.

General Term
Cloud Computing, Green Computing.

Keywords
Cloud computing environment, Green IT, resource allocation, Optimization.

1. INTRODUCTION
Cloud Computing is one of the mainly admired subject in the ICT sector nowadays. Cloud computing means incorporated, active infrastructures that carry IT as a service moreover inside (private cloud) or on the outside (public cloud). It is significant to recognize the trade-offs between Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), and among private and public clouds. Envision the potentials for the association if you may possibly persist to construct virtualized environment into an entirely mechanized, service- oriented transportation of collective resources (storage, server, and network) that permits to simply distribute IT Services to interior users. The cloud computing types is illustrated in fig 1.

There are at any rate three merits to optimizing the employment of a shared IT environment:

- Incredible agility
- Intense efficiency
- Highest exploitation

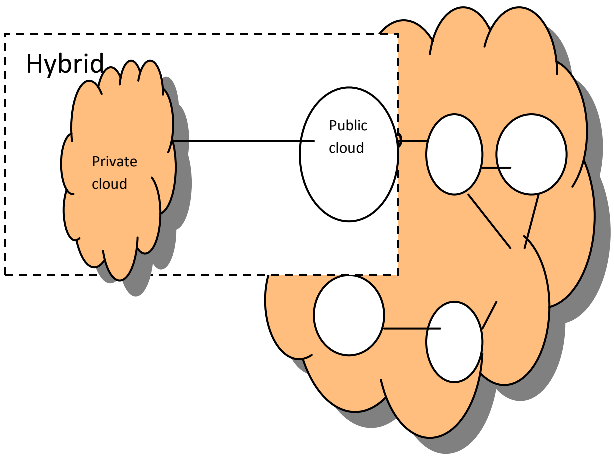


Fig 1: Cloud computing types

It combines cloud computing environment, cooling, saving power, and space, with money. Cloud presents a future-proof proposal that can develop no wildly as commerce requires. This indicates that you can revolve out novel applications sooner, be more receptive to client needs, and decrease IT costs on a huge scale by organizing a greatly competent infrastructure. The cloud computing construction is included of two considerable parts: the front end and the back end. The front end is the region at which the customer of the computer or the consumer himself is capable to access. The cloud computing outline might have predictable two principal apprehensions with the use of the cloud computing platform: privacy and security. Green Computing is particularly significant and appropriate: as computing develops into gradually more persistent, the energy consumption attributable to evaluating, regardless of the clarion identify to diminish utilization and turn around greenhouse effects. This is forcing the IT influential to hub on competence and total cost possession, predominantly in the framework of the world-wide economic crisis.

Environmental knowledge is the purpose of one or more of environmental science, environmental supervising, green chemistry, and electronic devices to scrutinize structure and defend the standard environment and resources, and to manage the destructive collides of human involvement. This is also phrased as green technology (abbreviated as *greentech*). The Green IT is also utilized to exemplify sustainable energy creation strategies such as Wind Turbine's, Photovoltaic etc. Sustainable progression is the central part of Green IT environment. The term *Green IT* is also illustrated a class of electronic devices that can encourage sustainable administration of resources.

In this work, we developed a resource distribution model of green cloud computing environments, by resending both processing capability and bandwidth are owed concurrently to every service request and borrowed out on an hourly basis. The billed resources are devoted to every service request and diminish the resource usage to make IT green.

2. LITERATURE REVIEW

Cloud computing is a representation for allowing expedient, on-demand system entry to a common pool of configurable calculating resources. Positioning and interior structures of resources have been studied [1], but power management is not done. Amazon Elastic Compute Cloud (EC2) [2] is an instance of HaaS (Hardware as a Service), which is a structure of cloud computing. Fairness should be followed whereas captivating numerous types of resource into deliberation. Cloud computing services are simple to use, and can decrease both trade costs and ecological loads [3]. Quick flexibility and measured service [4] are highlighted for cloud computing scenario. There are several papers that converse algorithms for attaining fairness for cases where a combined resource allocation is not measured [5]. To afford cloud computing services reasonably, it is significant to optimize resource distribution under the statement that the requisite fair resource [6] can be taken from a common resource group. Besides, to be able to present processing capability and storage ability, it is essential to assign bandwidth to entry them at the equal time. The paper [7] learns the optimization crisis of reducing resource leasing cost for managing flexible applications in cloud whilst gathering application service necessities. Such a crisis arises when unnecessary produced data acquires important economic cost on transmit and inventory in cloud. A monetary approach presented in [8], which services “offer” for possessions as a purpose of distributed performance.

Cloud computing services are quickly ahead in reputation. They permit the consumer to charge, only at the time when desirable, only a preferred quantity of calculating resources (processing capability and storage space capability) out of a massive distributed computing resources [9] without upsetting concerning the position or interior structures of these resources. The National Institute of Standards and Technology (NIST) recognized four necessary distinctiveness of cloud computing: resource pooling [10]. The reputation of cloud computing be obliged to amplify in the network speed, and to the reality that virtualization and network computing technologies have turn into commercially accessible. It is predictable that endeavors will hasten their movement from construction and possessing their individual systems to leasing cloud computing services [11].

To run common server resources [12], in this services “offer” for possessions as a purpose of distributed performance. In this work, an optimal resource allocation method is used for optimization of resource usage based on users’ task.

3. PROPOSED OPTIMAL RESOURCE ALLOCATION TECHNIQUE (ORAT) FOR GREEN CLOUD COMPUTING ENVIRORNMENT

The proposed work is efficiently designed to optimize resource allocation under the supposition that the requisite resource can be obtained from a joint resource pool. Besides, to be capable to offer processing capability and storage space facility, it is essential to assign bandwidth to entrée them at the similar time. The architecture diagram of the proposed Optimal resource allocation technique (ORAT) for green cloud computing is shown in fig 2.

As cloud computing services quickly enlarge their client support, it has developed into significant to afford them reasonably. To do so, it is necessary to optimize resource distribution under the statement that the necessary quantity of resource can be obtained from a widespread resource pool and borrowed out to the user on an hourly center. In addition, to be capable to present processing capability and storage space facility, it is required to preserve concurrently a network bandwidth to process them. Consequently, it is required to distribute several types of resources (such as processing capability, storage space capacity and bandwidth) concurrently in a synchronized way instead of assigning each type of resource separately. The quantity of resource vital and the period in which it is utilized are not permanent. They can differ really from user to user and since service to service.

The paper proposed an optimal resource allocation method for green cloud computing environments. For the preface assessment, this paper presumes the utilization of two types of resources: processing capability and bandwidth. In common, services can be divided into two groups: a non-delay system (defeat system) and a waiting scheme. A non-delay system assigns an additional resource instantly to the user ahead the advent of the request, and discards the request if there is no additional capacity. A waiting system assigns an additional capability to users in the series in which their needs have arrived, as a substitute of assigning resources instantly upon the advent of a request. This paper presumes a service that sprints as non-delay. This paper also thinks stationary resource distribution, which is the most essential structure of resource distribution, although active provision, which uses procedure immigration and bandwidth consolidation, can augment the exploitation of resources.

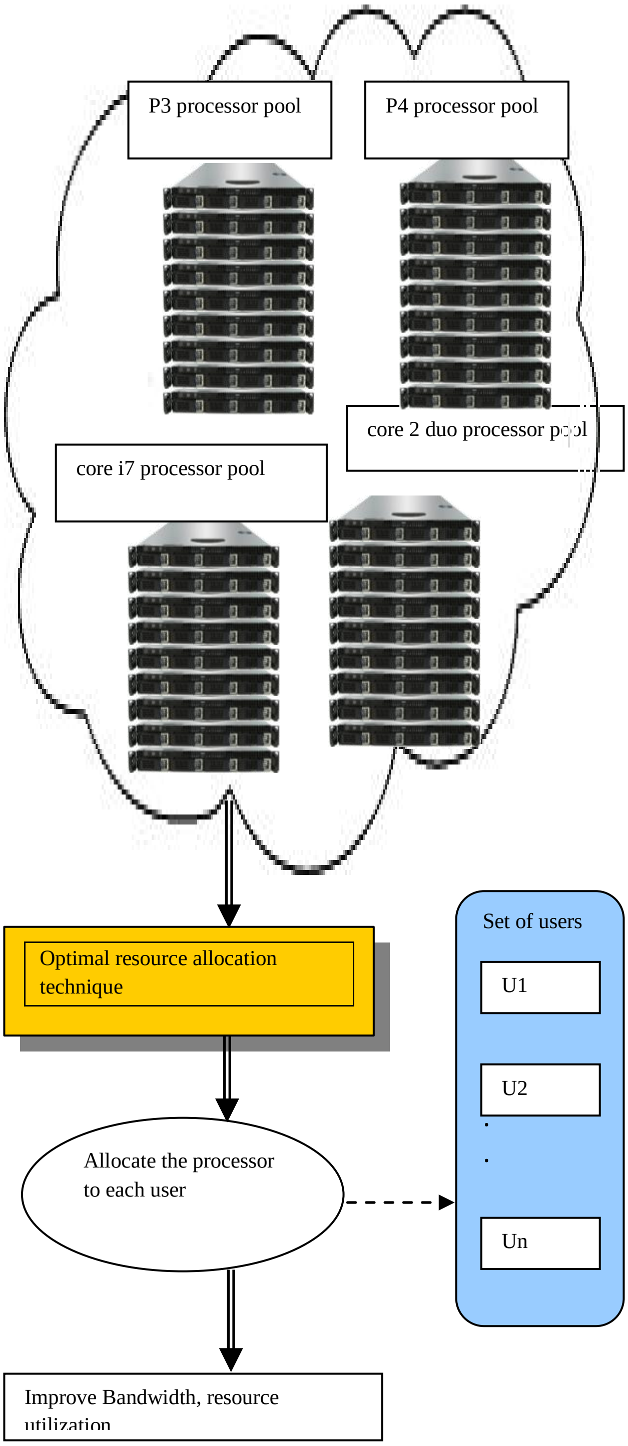
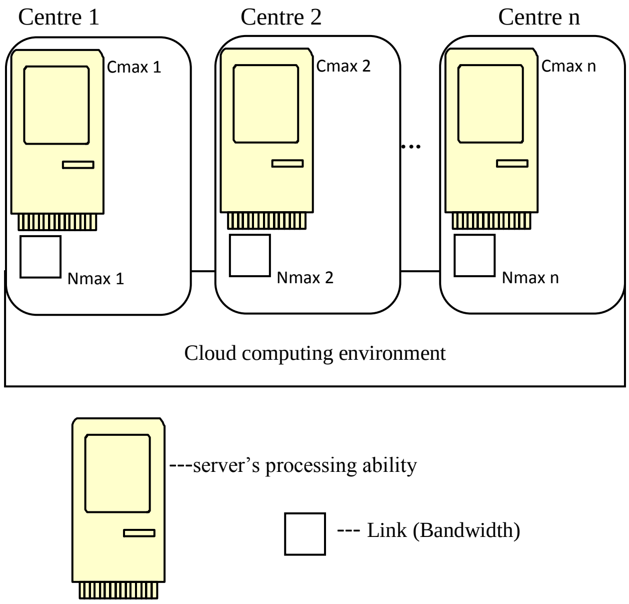


Fig 2: Architecture diagram of the proposed ORAT

3.1 Resource allocation in cloud computing environment

The resource allotment in a cloud computing environment can be represented as assigning the requisite quantity of numerous types of resource concurrently from a widespread resource pool for a definite stage of time for every request. The owed resources are committed (not shared) to every request. For the preface assessment, this paper proposes two types of resource: resource processing capability and bandwidth.

The requisite quantity of resource and the interlude of time in which it is utilized are not predetermined. They can differ significantly from user to user and from service to service. For instance, video delivery, file transfer, and videoconferencing services need a huge quantity of bandwidth but not so greatly resource processing capability. In distinction, a secretarial service needs a huge amount of processing capability but not so much bandwidth. It is understood that the hardware resources for green cloud computing services are not mounted at a distinct center, but in numerous biologically dispersed centers, as shown in Figure 3., in order to assist accumulation of resources, to employ load balancing and to guarantee high consistency. Each center has servers (counting virtual servers) that offer processing capability, and bandwidths that present access to these servers.



$C_{max\ j}$ - Maximum size of resource processing capability at center j
N_{maxj} - Maximum size of bandwidth at center j

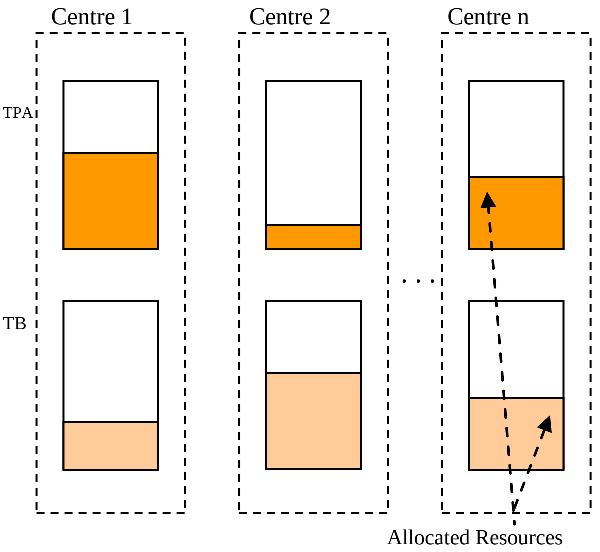
Fig 3: System model for cloud computing services

When a new request is produced, one hub from between k centers is chosen consistent with the resource distribution algorithm. The highest size of processing capability and bandwidth at center j (j=1, 2,..,n) is specified to be C_{maxj} and N_{maxj} correspondingly. The notion of resource allocation obtains the resource practice stage into deliberation. When a service request appears, the finest center is chosen from numerous centers, and both the processing capacity and bandwidth accessible in the chosen center are owed for a definite period of time. The processing capacity and bandwidth of only this particular center are owed. If no center has a sufficient quantity of auxiliary resources (both processing capacity and bandwidth), the request is discarded. The resources owed are unconfined after the usage period has gone.

3.2 Optimal resource allocation technique for green cloud computing environment

The objective of the proposed optimal resource allocation technique (ORAT) for green cloud computing allocation is to exploit the number of requests to which both resource processing capability and bandwidth are maintained well. As the amount of requisite resource processing capability does not usually have a rigid association with that of requisite bandwidth, the finest resource allocation cannot be attained if only a single resource kind is measured in the assortment of a center.

The resource that needs the prevalent balanced size of resource, contrast the amount of requisite resource with the greatest resource size for every resource type, is first chosen as 'recognized resource'. Then the smallest amount accessible of the recognized resource from amongst k centers is chosen. In this case, the parts of resource processing capability and bandwidth are diverse, being considered in proportion of CPU power and b/s (bits per second) correspondingly. It is proposed to evaluate the size of the diverse resources as in the subsequent instance. Suppose that the highest quantity of bandwidth in a center is 100Mb/s. A demand for 20% of CPU power and 30Mb/s needs 20% of resource processing capacity and 30% of bandwidth correspondingly. As the amount of requisite bandwidth is better than that of requisite processing capability, bandwidth will be the known resource in this case.



TPA-Total Processing Ability
TB-Total Bandwidth

Fig 4: Possible resource allocation for a user

When a service request task is generated by the user, the subsequent resource allocation algorithm is conceded out and shown in fig 4.

i) Assortment of known resource

if $X_C > X_N$ then processing capacity is the recognized resource.

Else bandwidth is the recognized resource.

Where

X_C = {the amount of necessary processing capacity}/XC0

X_{C0} = Min {the highest amount of processing facility in a center}

X_N = {the size of requisite bandwidth}/XN0

X_{N0} = Min {the greatest amount of bandwidth in a center}

For instance, if there are two centers and the greatest amount of resource processing ability of every center is 100 and 50 correspondingly, XC0 will be 50.

ii) Assortment of the center

The center which suits the subsequent three conditions will be chosen:

a.) Min {the obtainable size of the notorious resource in the center}

b.)Accessible processing capacity in the center is equivalent to or superior than the requisite processing ability.

c.) Accessible bandwidth in the center is equivalent to or better than the requisite bandwidth.

If there are two or more centers which persuades the above conditions, one center will be chosen at arbitrary. Remind that the request will be discarded if there are no centers that persuade the above conditions.

iii) Distribution of resource

Both requisite processing capacity and bandwidth are owed concurrently in the chosen center. The maintained resources are applied (not common) to the request.

iv) Resource released

When the service time has concluded, both owed processing capability and bandwidth will be confined concurrently.

The above algorithm is followed for resource allocation process based on an optimization of resource usage based on CPU cycles, bandwidth, processing ability it needs to process the user given tasks and an experimental evaluation is conducted to estimate the performance of the proposed ORAT and described in next section.

4. EXPERIMENTAL EVALUATION

The proposed Optimal resource allocation technique (ORAT) for green cloud computing is implemented in Java using cloudsim software. The proposed optimal resource allocation technique (ORAT used old range, mid range and high end processors. The old range processors includes 286, 386, Pentium, the mid range processors includes Pentium pro, Pentium III, Pentium IV, the high end processors be core2, core i7. These types of processors pools are integrated to analyze the performance of the proposed optimal resource allocation technique (ORAT) for green cloud computing based on processors utilization. The number of tasks assigned to the processor is based on the capability of the processor in the resource pool measured in terms of CPU cycles, bandwidth, and data rate. The proposed ORAT first identifies the tasks schedule, resource/processor capability. An optimized resource allocation takes place under the assumption that the required resource can be taken from a shared resource pool. In addition, to be able to provide processing ability and storage capacity, it is necessary to allocate bandwidth to access them at the same time. Operations can be assigned to the pool of old range and mid range processors with high end processors. The proposed ORAT for cloud computing infrastructure is measured in terms of:

i.)Request loss probability.
ii.)Resource utilization.
iii) Data transfer rate.

5. RESULTS AND DISCUSSION

In this work, we have seen how a pool of processors can be allocated to users' tasks based on optimal resource utilization technique to capture the performance of cloud computing process in Green IT to other systems written in mainstream languages such as Java. We run independent tests with several number of resources task with a constant number of tasks sent by each users. The entire process of the proposed Optimal resource allocation technique (ORAT) for green cloud computing is explained in section 3 briefly and this section described the performance of the proposed architecture. Compared with an existing integrated time and task based process schedule to reduce the accumulate the e-waste which describes only about the assignment of users' task schedule not keen about the resource wastage, the proposed ORAT architecture provides a better results in terms of utilization of resource for eradicating the e-waste by accumulating minimum amount of energy. The given below table and graph shows the result of the proposed ORAT for green cloud computing compared with existing integrated time and task based process schedule.

Table 1. No. of Resources Vs. Request loss probability

No. of requests	Request loss probability (%)	
	Proposed ORAT for green cloud computing	Existing ITTPS
2	0.5	0.8
4	0.9	1.6
6	1.2	1.9
8	1.4	2.3
10	1.5	2.5

The above table (table 1) describes the users' task request loss probability to determine an efficient resource utilization. The effect of the proposed Optimal resource allocation technique (ORAT) for green cloud computing is compared with an existing integrated time and task based process schedule.

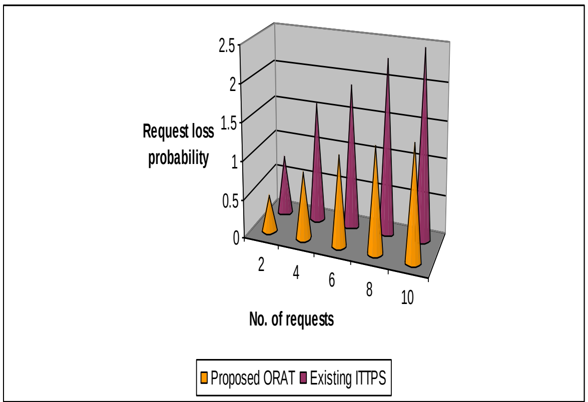


Fig 5. No. of Resources Vs. Request loss probability

Fig 5. describes the users' task request loss probability based on number of resources present in the cloud computing environment. It compares the request loss prospect in the situation where the sizes of processing capability and bandwidth (C and N) ascend and drop in anti-phase, i.e., a great processing capability is pursued by a huge bandwidth. Assess the impact of ratio of greatest resource size of every center on the request loss possibility, assuming the total size of processing capability (Cmax1+Cmax2) and that of bandwidth (Nmax1+Nmax2) are stable. The proposed ORAT can decrease the request loss possibility and as a effect, decrease the total amount of resource contrast with an existing integrated time and task based process schedule, in the situation where the sizes of processing capacity and bandwidth increase and descend in anti-phase. This is also accurate even if the number of centers augments apart from for the case where the number of centers odd.

Table 2. No. of Users Vs. Average Resource Utilization

No. of users	Average resource utilization (%)	
	Proposed ORAT for green cloud computing	Existing ITTPS
5	54	31
10	63	43
15	75	52
20	82	60
25	89	65

The above table (table 2.) describes the utilization of resources in a meaningful manner without producing e-waste to determine efficient resource utilization. The effect of the proposed Optimal resource allocation technique (ORAT) for green cloud computing is compared with an existing integrated time and task based process schedule.

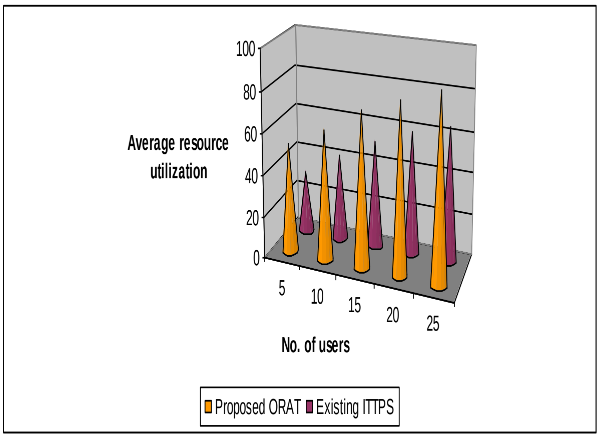


Fig 6. No. of Users Vs. Average resource utilization

Fig 6. describes the optimal resource utilization based on number of users present in the cloud computing environment with the tasks they need to process. In the proposed ORAT, the resource allocation is made to the users' task based on the CPU cycles, bandwidth and data rate of the respective resource. Based on the resource capability, the task has been assigned by the framework to evaluate the utilization of resource in a meaningful manner without throwing as a waste. The resource utilization is measured in terms of average number of resources are entirely used by the tasks allotted to it. Compared to an existing ITTPS, the proposed optimal resource allocation technique (ORAT) for green cloud computing outperforms well in resource utilization and the variance is 40-50% high in the proposed ORAT.

The below table (table 3.) describes the transfer rate of data based on number of tasks hold by the users. The effect of the proposed Optimal resource allocation technique (ORAT) for green cloud computing is compared with an existing integrated time and task based process schedule.

Table 3. No. of tasks Vs. Data transfer rate

No. of tasks	Data transfer rate	
	Proposed ORAT for green cloud computing	Existing ITTPS
4	45	19
8	58	24
12	64	33
16	71	39
20	80	43

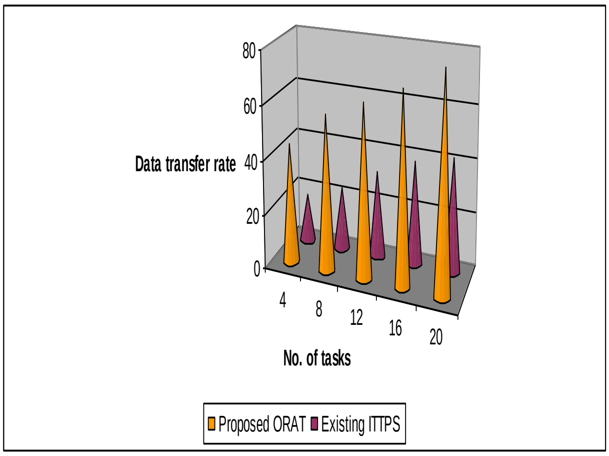


Fig 7. No. of tasks vs. Data transfer rate

Fig 7. describes the transfer rate of data based on number of tasks hold by the users present in it. In the proposed ORAT, the data transfer rate is high, since it followed the resource utilization path by adapting the optimal resource utilization technique. The data transfer rate is measured in terms of kilobytes per second (kbps). Compared to an existing ITTPS, the proposed optimal resource allocation technique (ORAT) for green cloud computing performs a reliable data transfer rate in resource utilization and the variance is 45-50% high in the proposed ORAT.

At last, it is observed that the proposed optimal resource allocation technique (ORAT) for green cloud computing outperforms well by adapting the optimal resource allocation method. The proposed ORAT for cloud computing assigned the processors to the tasks based on its Resource Usage (bandwidth, CPU cycles, I/O operation, Energy, and data transfer rate). Based on the resource capabilit, the task has been assigned to the resource, so it consumes less time to process and execuet the task and it utilize all the resources without any processors waste.

6. CONCLUSION

The proposed paper first has presented a resource allocation model for green cloud computing environments and has proposed the optimal resource allocation method, considering that both resource processing capacity and bandwidth are distributed concurrently for each request and returned out on an hourly basis. The distributed resources are assigned to each service request. It has been established by simulation evaluation that the proposed optimal resource allocation technique (ORAT) could decrease the request loss possibility and as a consequence, decrease the total sum of resource utilized, contrast with an existing ITTPS. Then, this paper has proposed basic measures for attaining fair resource distribution amongst several users in a cloud computing environment, which endeavors to assign resources in part to the estimated quantity of resources demanded by each user. The proposed optimal resource allocation technique (ORAT) enables the resource allocation amongst multiple users without a huge beg off in resource efficiency, contrast with an existing ITTPS method which does not judge the fair allocation of resources.

7. REFERENCES

[1] G.Reese, Apr. 2009.*"Cloud Application Architecture"*, O'Reilly & Associates, Inc.

[2] J.W.Rittinghouse and J.F.Ransone, Aug. 2009. *"Cloud Computing: Imprementation, Management, and Security"*, CRC Press LLC.

[3] P.Mell and T.Grance, July 2009. *"Effectively and securely Using the Cloud Computing Paradigm"*, NIST, Information TechnologyLab.

[4] P.Mell and T.Grance, 2009. *"The NIST Definition of Cloud Computing"*, Version 15.

[5] P.Yue, Z.Liu and Z.Zhang, April 2006. *"Fair bandwidth allocation for responsive and unresponsive flows using approximate fairness dropping scheme"*, IEICE Trans. Commun., Vol.E89-B, No.4.

[6] K.Hatakeyama, S.Tsumura and S.Kuribayashi.Oct.2008. *"Fair joint multiple resource allocation method in all-IP networks"*, In Proc. of 14th Asia-Pacific Conference on Communications (APCC2008), 15-PM1-F-3.

[7] Han Zhao ET.AL., 2012. *"Optimal Resource Rental Planning for Elastic Applications in Cloud Market"*, Proceedings of 26th IEEE International Parallel and Distributed Processing Symposium, IPDPS.

[8] Jeffrey S. Chase, Darrell C. Anderson, 2001.*"Managing Energy and Server Resources in Hosting Centers"*, Proceedings of the eighteenth ACM symposium on Operating systems principles.

[9] J. Zhang, J. Kim, M. Yousif, R. Carpenter, and R. J. Figueiredo, 2007.*"System-level performance phase characterization for on-demand resource provisioning,"* *Proceedings of the 2007 IEEE International Conference on Cluster Computing (CLUSTER '07)*, pp. 434–439.

[10] P. Padala, K. G. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal, A. Merchant, and K. Salem, March 2007.*"Adaptive control of virtualized resources in utility computing environments,"* *SIGOPS Oper. Syst. Rev.*, vol. 41, pp. 289–302.

[11] Z. Gong, X. Gu, and J. Wilkes,2010. *"Press: Predictive elastic resource scaling for cloud systems"*, *International Conference on Network and Service Management (CNSM '10)*, pp. 9–16.

[12] H. M. Monti, A. R. Butt, and S. S. Vazhkudai, 2011. *"Catch: A cloud-based adaptive data transfer service for hpc,", IEEE International Parallel Distributed Processing Symposium (IPDPS '11)*, pp. 1242–1253.

8. AUTHOR'S PROFILE

K.L.Giridas received his master's degree in computer science and engineering from PSNA College of engineering and technology, India, 2005. He worked as Professor and Head of the department of faculty of Information Technology. He is currently a PhD student in the Faculty of Computer Science and Engineering at Noorul Islam University (NIU), Tamilnadu, India. His research interests include cloud computing, green computing, and computer architecture.

A.Shajin Nargunam is a professor of computer science and engineering at the Noorul Islam University (NIU), Tamilnadu, in India. He is now Academic Director of Noorul Islam University (NIU) at Noorul Islam Centre for Higher Education (NICHE). He received his PhD in computer science and engineering from National Institute of Technology (NIT), Calicut in 2009. In 2003, he won the appreciation award for designing a low cost super computer at Science and technology exhibition. Shajin had been working at NIU for the past fifteen years, since 1997. He has got so many international and national journal publications into his credit. His research interests include mobile computing, cloud computing, virtualization technology, cluster computing, peer-to-peer computing, network storage and network security.